

J_{ET}^{II} : A Two-prong Jet Finding Algorithm

Yang Bai,^a Zhenyu Han^b and Ran Lu^a

^a *Department of Physics, University of Wisconsin-Madison, Madison, WI 53706, USA*

^b *Institute for Theoretical Science, University of Oregon, Eugene, OR 97403, USA*

We propose a new global jet-finding algorithm for reconstructing two-prong objects like hadronic weak gauge bosons at a hadron collider. The selection of particles in a two-prong jet is required to maximize a J_{ET}^{II} function, which contains a modified second Fox-Wolfram moment and prefers a two-prong structure for a fixed jet mass. Compared to the traditional jet-substructure method, our algorithm can provide a similar or better performance for identifying boosted weak gauge bosons that are produced from either Standard Model processes or heavy di-boson resonance decays.

Introduction. Jet substructure techniques have become essential tools in the search for new physics at a hadron collider such as the Larger Hadron Collider (LHC). Strategies have been proposed to reconstruct hadronically decaying objects such as the W and Z gauge bosons, the Higgs boson and the top quark, and extract from them vital information that was perviously unattainable. Most existing jet substructure methods are separated into two steps: first, jets are formed with standard jet-finding algorithms [1, 2]. The jet size must be large enough to include most of the decay products of the object. Therefore the jets are “fat”. Fat jets are then processed with a grooming algorithm [3–7] or other methods to remove soft particles that are unlikely from the object decay. An alternative, relatively unexplored approach is to develop a new jet-clustering algorithm to directly reconstruct the multi-prong objects, utilizing its structure and properties.¹ The algorithm could be more specialized to suit different objects, and, if carefully designed, should be more efficient to reconstruct the physical object than ordinary jet-finding algorithms because it contains more physics information.

To reconstruct a single-prong jet, one can choose a set of particles to maximize a jet function that prefers a large jet momentum but a small jet mass [11–14].² The relevant jet function for one-prong jet only depends on the summed four-momenta of particles belonging to the jet. Following the same philosophy, in this paper we develop a new jet-finding algorithm to reconstruct two-prong jets. Since the two-prong structure can be interpreted as a jet shape, the immediate question is how to add an additional component in the jet function to prefer a two-prong structure. There is a large amount of literature on event shapes at a lepton collider including the Fox-Wolfram (FW) moments [16, 17]. We will show that

with some modifications, we can incorporate the second Fox-Wolfram moment into the jet function, maximizing which prefers a two-prong structure for a jet.

Our new jet-finding algorithm is a global one because all subsets of particles will be considered to maximize the jet function. It has advantages of including only relevant particles for the physical object, and being infrared and collinear safe based on the argument in Ref. [11] and the discussion on FW moments in Refs. [16, 17]. On the other hand, it also has its disadvantage because 2^N subsets need to be checked, where N is the number of particles in an event, and an unrealistic computing power is required for just dozen of particles. To reconcile this problem, we provide a double-anti- k_t -cone algorithm to find the approximate maximum of the jet function. As we will demonstrate, this approximation matches very well the true global maximum. As a result, the double-anti- k_t -cone algorithm has a realistic number of operations of $\mathcal{O}(N^2 n_R^3)$, where n_R is the number of different anti- k_t jet sizes.

Two-prong Jet Function. Similarly to the studies in Refs. [11–13] for a one-prong jet, we maximize the following two-prong jet function to select the subset of particles in an event as a jet

$$J_E^{\text{II}} = E^2 - \beta M^2 + \gamma \tilde{H}_2 \quad (\text{lepton colliders}), \quad (1)$$

$$J_{ET}^{\text{II}} = E_T^2 - \beta M^2 + \gamma \tilde{H}_2 \quad (\text{hadron colliders}). \quad (2)$$

Here, E and M are the jet energy in the lab frame and jet mass for the subset of (massless) particles. For hadron colliders, we use the transverse energy defined as $E_T^2 \equiv E^2 - P_z^2$. In addition to the one-prong jet function [13], we have added an additional component,

$$\tilde{H}_2 = M^2 \left[\sum_{i,j} \frac{(p_i \cdot p_j)^2}{(P_J \cdot p_i)(P_J \cdot p_j)} - 1 \right], \quad (3)$$

with $P_J^\mu \equiv \sum_i p_i^\mu$ the total four-momentum of the subset and $p_i \cdot p_j = p_i^\mu p_{j\mu}$.

A simpler way to understand this new function is to boost all particles to the rest frame of P_J^μ , in which one

¹ The N -jettiness/XCone [8–10] method can be viewed as such an attempt, although in which a fixed number of jets in an event are constructed.

² See Ref. [15] for its connection to other jet finding algorithms.

has

$$\tilde{H}_2 \equiv \left(\sum_{i,j} E_i E_j \cos^2 \theta_{ij} \right)_{\text{rest frame}}. \quad (4)$$

Here, θ_{ij} is the angle between two particles in the rest frame of P_J^μ . For a fixed invariant mass $\sum_i E_i^{\text{rest frame}} = M$, one can easily show that \tilde{H}_2 is maximized when there are only two back-to-back massless prongs. Our function is related to the FW moments [16, 17], $H_l \equiv \sum_{i,j} |\vec{p}_i| |\vec{p}_j| P_l(\cos \theta_{ij})/E^2$, by $\tilde{H}_2 = E^2(2H_2 + 1)/3$. It is also interesting to note that the zeroth and first FW moments are equivalent to the jet energy and mass, respectively. Therefore, our J_E^{II} can be expressed just using $H_{0,1,2}$.

For the one-prong jet function in Refs. [11–13], it was shown that the jet parameter β controls the size of the jet with the jet cone radius scaling as $1/\sqrt{\beta}$ for a large β . Our two-prong jet functions in Eqs. (1) and (2) have two parameters, β and γ . Similar to the one-prong case, β should be positive to avoid clustering all particles in an event into a single jet. Furthermore, we also require $\beta \geq \gamma$, otherwise two back-to-back groups of particles are preferred in a single jet. This can be easily seen for two massless energetic particles, which have $J_E^{\text{II}} = E^2 - (\beta - \gamma) M^2$ and $(\beta - \gamma)$ as the effective β for the one-prong jet function. Therefore, we anticipate that the parameter $1/\sqrt{\beta - \gamma}$ for a large $(\beta - \gamma)$ roughly controls the geometric size of a fat two-prong jet. At hadron colliders, in order to exclude particles in the beam direction, we also require $\beta - \gamma > 1$.

To further understand the shape of an object obtained by maximizing the two-prong jet function, we add an infinitely soft particle into the system of two energetic massless prongs to determine the passive catchment area [18]. Since the generic feature at a hadron collider is similar to the one at a lepton collider, we study the jet function at a lepton collider for simplicity. Defining the weighted angular distances of the soft particle with respect to the hard particles

$$d_1 = \frac{2 E_1 (1 - \cos \theta_1)}{E_1 + E_2}, \quad d_2 = \frac{2 E_2 (1 - \cos \theta_2)}{E_1 + E_2}, \quad (5)$$

with θ_1 and θ_2 as the angles between the soft particle and the two hard particles, we derive the following condition for the soft particle to increase the value of J_E^{II}

$$\beta(d_1 + d_2)^2 - \gamma(d_1 - d_2)^2 < 2(d_1 + d_2). \quad (6)$$

Choosing the two hard particles with $E_1 = 3E_2$ located at $(\pi/2, \pm\pi/12)$, we show the boundary of soft particles in Fig. 1 for different values of β and γ . From Eq. (6) and Fig. 1, one can see that for a large value of β the soft particles are close to one of the two hard particles and the allowed region consists of two separated cones. As β decreases, the two cones expand and overlap with

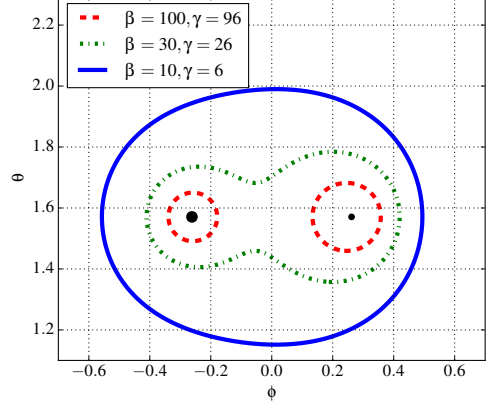


FIG. 1: The boundaries of soft particles surrounding two hard particles with $E_1 = 3E_2$ on the $\theta - \phi$ plane for different values of β and γ . The locations of the two hard particles are marked with the black dots.

each other. For a small value of β , the “fat jet” has a single cone with two subjets inside, which is what the traditional jet-substructure method searches for.

Double-anti- k_T -Cone Algorithm. It is unrealistic to exhaust all 2^N possible subsets of particles to maximize our jet function, even for several dozen of particles in an event. In practice, however, one can find approximate and fast ways to identify a subset of particles that provide a jet function value close to the true maximum. We provide one such algorithm built on top of the anti- k_t algorithm [19].

Given the observation in Fig. 1 that the jet shape can be described roughly by two separate or overlapping cones, one could consider all possible pairs of cones (and individual ones), and identify the one with the largest $J_{E_T}^{\text{II}}$ as the jet candidate. Specifically, our jet-finding algorithm is described as following

1. Preparing candidate cones: for a given event, collect anti- k_t (or other fast cone-shape-jet-finding algorithm) jets with various jet radii. For instance, one can scan the jet radii in $n_R \sim \mathcal{O}(10)$ steps with a maximum value of $R_{\text{max}} \sim \mathcal{O}(1)$. In the following simulation, we will choose $R_{\text{max}} = 0.7$ and $R_{\text{min}} = R_{\text{max}}/n_R = 0.05$.
2. Maximizing $J_{E_T}^{\text{II}}$: for each pair of (and individual) anti- k_t jets, calculate $J_{E_T}^{\text{II}}$ based on the particles belong to this pair of (or individual) anti- k_t jets. Particles shared by two different jets are counted only once. Find the pair (or the individual one) that maximizes $J_{E_T}^{\text{II}}$ as the leading jet.
3. Iterating: remove particles that belong to the leading jet, and repeat the procedure to find the next jet, until all particles in an event are exhausted.

The running time of this algorithm for N particles scales as $N^2 n_R^3$.³ We test the goodness of the above approximate algorithm against the true maximum using a sample of hadronic WW events generated with a $p_T^W > 200$ GeV cut. Since the true maximum requires us to exhaust all 2^N possible subsets of particles, we only keep the leading 20 particles in our test. We have found that over 99% of the events, the double-anti- k_T -cone algorithm finds exactly the same set of particles for the leading jet as the exhaustive subset method. For the remaining less than one percent of events, the particle contents only differ slightly with the difference of the jet p_T 's less than 10%.

Results for W -jet Tagging. Using boosted W -bosons as an example, we compare the results from our J_{ET}^{II} jet-finding algorithm with other traditional jet-substructure methods including the jet grooming methods [3–6]. For the signal, we consider the Standard Model WW pair production at the 14 TeV LHC with both W 's decay hadronically. The dominant background is QCD dijets. Both signal and background events are simulated with PYTHIA 8 [21]. To mimic the pileup effects, we also add 20 soft QCD events for each hard event.

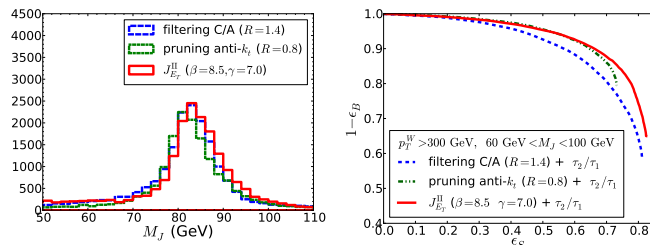


FIG. 2: Left panel: jet mass distributions for the leading two jets. The jet algorithm parameters for obtaining the largest signal efficiency can be found in Table I for $p_T^W > 300$ GeV. Right panel: the receiver operating characteristic curves for the signal and background efficiencies by varying the cut on the N -subjettiness ratio variable τ_2/τ_1 [22].

Imposing a parton-level $p_T > 300$ GeV cut on the W 's and QCD jets, we show the reconstructed jet mass distributions in the left panel of Fig. 2, comparing our jet-finding algorithm and the traditional jet-substructure analysis with a fat jet plus filtering or pruning. We have chosen to maximize the signal acceptance for jet mass in the (60,100) GeV window, and the optimum jet parameters are shown in Table I. From the left-panel of Fig. 2, one can see that our jet-finding algorithm identifies more W -jets in the mass window. For events within the (60,100) GeV mass window, we further calculate the N -subjettiness ratio, τ_2/τ_1 [22]. We apply a varying cut on τ_2/τ_1 to produce the receiver operating characteristic

(ROC) curves, shown in the right panel of Fig. 2. Comparing to filtering and pruning, our jet algorithm provides a comparable or larger signal acceptance for most of the background efficiencies. The ROC curve of pruning is very close to the J_{ET}^{II} one when the signal efficiency is below 60%.

$p_T^W >$	Filtering		Pruning		J_{ET}^{II}	
	$R(\text{C/A})$	R_{filter}	$R(\text{anti-}k_t)$	z_{cut}	β	γ
200 GeV	1.6	0.35	1.0	0.075	6	5
300 GeV	1.4	0.30	0.8	0.05	8.5	7
400 GeV	1.0	0.25	0.8	0.05	11	9.5
800 GeV	0.6	0.15	0.6	0.025	12	7

TABLE I: The optimized jet-finding parameters that provide the maximal signal efficiency for different p_T cuts in the jet mass window of (60,100) GeV. For pruning, we start from an anti- k_t fat jet and prune it using the Cambridge/Aachen (C/A) algorithm. The additional parameter D_{cut} is fixed to be $D_{\text{cut}} = m_J/p_{TJ}$. A smaller value of D_{cut} provides marginally larger signal efficiencies, but with the signal peak shifted to a lower value.

To test our jet-finding algorithm for W 's with different p_T 's, we scan the jet parameters to obtain the maximal signal efficiency and show the values of the jet parameters in Table I. For the fat jet plus pruning method, we choose to start with anti- k_t fat jets, which give us slightly better results than starting from Cambridge/Aachen (C/A) [23, 24] fat jets, and then prune them using the C/A algorithm. For the fat jet plus filtering method, we use the C/A algorithm for both the fat jet and the filtering procedure. We have found that including only two subjets provides a larger signal efficiency in the jet mass window of (60,100) GeV, than three subjets. For our J_{ET}^{II} algorithm, larger values of β and $\beta - \gamma$ are preferred for W 's with a larger p_T .

After fixing the jet parameters for different p_T cuts, we show the background fake rate in Fig. 3 with a fixed signal efficiency of 70%. From Fig. 3, one can see that for less boosted W 's with a small p_T cut, the filtering method better identifies the two isolated jets and does a slightly better job in rejecting the background. For very boosted W 's, our J_{ET}^{II} and the pruning methods provide comparable results, which are better than the filtering method because they do not force two separate prongs as the filtering does. For the intermediate p_T cuts, our J_{ET}^{II} algorithm has smaller background fake rates than the other two methods. Therefore, one can use our method to more efficiently tag hadronic W -jets for a wide range of p_T 's.

Results for Heavy Diboson Resonance. To further illustrate the performance of our algorithm, we consider a beyond-the-Standard-Model-physics example, motivated by the recent ATLAS searches [25] for diboson resonances using jet-substructure techniques. We consider a heavy

³ Our numerical code can be found in <https://github.com/LHCJet/JETII>, which is based on FastJet [20].

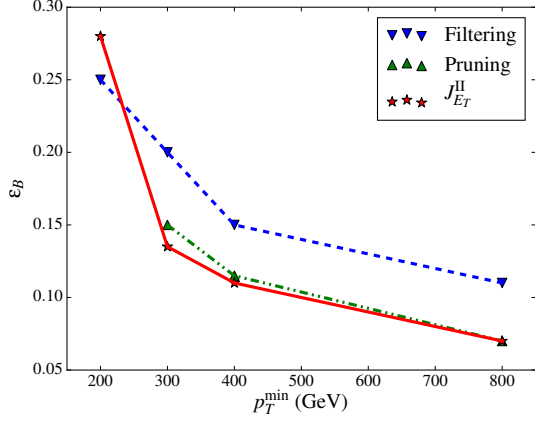


FIG. 3: The QCD dijet background fake rate for different p_T cuts with a fixed signal efficiency at 70%. The jet mass window is chosen to be (60, 100) GeV. The pruning case with $p_T > p_T^{\min} = 200$ GeV is not shown because the maximal signal efficiency that it can reach is only around 60%.

W' gauge boson in the model of Ref. [26], produced at the 8 TeV LHC with 20 fb^{-1} integrated luminosity. We fix the W' mass to be 2 TeV and let it decay to $W + Z$, both of which decay hadronically.

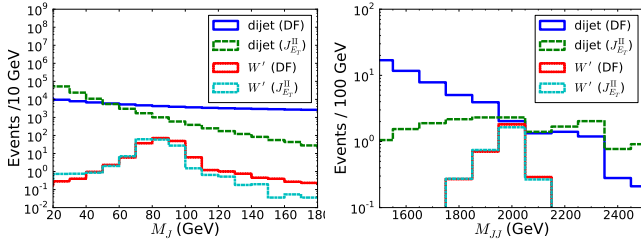


FIG. 4: Left panel: the signal and background distributions for the leading two jet masses from our $J_{E_T}^{\text{II}}$ and declustering-filtering (DF) procedures. Right panel: the distributions for the invariant masses of two leading jets with the lower (higher) jet mass within a 26 GeV mass window centered at the $W(Z)$ mass.

Following the ATLAS analysis in Ref. [25], we also use the declustering-filtering (DF) procedure to identify the W and Z bosons and to keep the QCD dijet mistag rate to be around 1%. Starting from $R = 1.2$ C/A fat jet with $p_T > 540$ GeV, we impose the same cuts on the mass-drop, subjet momentum balance and filtering parameters with $\mu_f = 1$, $\sqrt{y_f} = 0.2$ and $R_r = 0.3$.⁴ For our $J_{E_T}^{\text{II}}$ jet-finding algorithm ($\beta = 100$ and $\gamma = 92$) and to further

suppress the QCD background, we impose an additional cut on the momentum balance to utilize the fact that a QCD jet has more radiation around its axis than a W/Z jet: in each fat jet, we construct the subleading $J_{E_T}^{\text{II}}$ jet around the leading one and require its transverse momentum to be smaller than 0.015 times that of the leading one.⁵

After selecting the two leading jets with the lower (higher) jet mass within a 26 GeV mass window centered at the $W(Z)$ mass, we require them to have $|y_1 - y_2| < 1.2$ and $(p_{T1} - p_{T2})/(p_{T1} + p_{T2}) < 0.15$ like in the ATLAS analysis. We show the distributions of the individual jet mass in the left panel of Fig. 4 and the dijet invariant mass distributions in the right panel. The left panel of Fig. 4 clearly shows that our $J_{E_T}^{\text{II}}$ algorithm is more efficient than the DF method to reduce QCD dijet events with a jet mass within the W/Z boson mass window. From the right panel and imposing an invariant mass window cut of (1.8, 2.2) TeV, our $J_{E_T}^{\text{II}}$ method has improved S/\sqrt{B} by 15% compared to the DF method. We also note here that the QCD dijet invariant mass distributions from our method has a harder tail and a bump around 1.9 TeV. This can be understood from the basic properties of our jet function in Eq. (2), which prefers jets with a small jet mass but a large p_T . The bump location is generated from the basic p_T cut on the leading two jets and the additional cut on the surrounding subleading jet p_T .

Discussion and Summary. Our jet-finding algorithm has the advantage of identifying a smaller set of particles relevant for the boosted two-prong objects. On the other hand, it cannot capture other properties of boosted objects like the number of charged particles, the momentum balance of subjets or other color information of the jet [28, 29]. One should also apply the existing jet-substructure variables to our $J_{E_T}^{\text{II}}$ jets to further enhance the new physics discovery potential.

One could simply apply the $J_{E_T}^{\text{II}}$ algorithm to identify the boosted Higgs boson with $h \rightarrow b\bar{b}$. For the single b -tagging case, one can first identify a two-prong jet and then check the additional b -tagging information for particles belong to this jet. For the double b -tagging case, $h \rightarrow b\bar{b}$ already guarantees the two-prong structure and our $J_{E_T}^{\text{II}}$ may not add any additional discrimination power. It is interesting to see how to extend our $J_{E_T}^{\text{II}}$ algorithm together with the b -tagging method.

Beyond the two-prong jets, one could follow the general philosophy in this paper to identify a jet function favoring other decay topologies. There is not a single Fox-Wolfram moment that favors a three-prong jet like a hadronic top. On the other hand, it is still possible

⁴ Our goal is not to exactly reproduce ATLAS results, which contain a detector simulation. For instance, we do not impose the number-of-track cut in our analysis, which could be more sensitive to detailed pileup information.

⁵ One could also design a “telescoping” jet cut [27] by choosing a different set of β and γ .

to define a jet function using a series of Fox-Wolfram moments, or other functions constructed to match the special event shape [30].

In conclusion, we have introduced a new jet-finding algorithm that favors two-prong jets. For the Standard Model WW productions with two hadronic W 's, our $J_{E_T}^{\text{II}}$ algorithm provides performance close to or better than the results from traditional jet grooming methods like filtering and pruning for a wide range of p_T 's. For a heavy di-boson resonance decaying to two hadronic weak gauge bosons, our $J_{E_T}^{\text{II}}$ algorithm can further improve the new particle discovery significance.

Acknowledgments. We would like to thank Vernon Barger, Steve Ellis, Jesse Thaler and Lian-Tao Wang for useful discussion and comments. Y. Bai and R. Lu are supported by the U. S. Department of Energy under the contract DE-FG-02-95ER40896. Z. Han is in part supported by U. S. Department of Energy under grant numbers DE-FG02-96ER40969 and DE-FG02-13ER41986.

-
- [1] S. D. Ellis and D. E. Soper, Phys. Rev. **D48**, 3160 (1993), arXiv:hep-ph/9305266 [hep-ph] .
 - [2] G. P. Salam, Eur. Phys. J. **C67**, 637 (2010), arXiv:0906.1833 [hep-ph] .
 - [3] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys.Rev.Lett. **100**, 242001 (2008), arXiv:0802.2470 [hep-ph] .
 - [4] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Phys. Rev. **D80**, 051501 (2009), arXiv:0903.5081 [hep-ph] .
 - [5] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Phys. Rev. **D81**, 094023 (2010), arXiv:0912.0033 [hep-ph] .
 - [6] D. Krohn, J. Thaler, and L.-T. Wang, JHEP **02**, 084 (2010), arXiv:0912.1342 [hep-ph] .
 - [7] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, JHEP **05**, 146 (2014), arXiv:1402.2657 [hep-ph] .
 - [8] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, Phys. Rev. Lett. **105**, 092002 (2010), arXiv:1004.2489 [hep-ph] .
 - [9] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion, and T. F. Wilkason, (2015), arXiv:1508.01516 [hep-ph] .
 - [10] J. Thaler and T. F. Wilkason, (2015), arXiv:1508.01518 [hep-ph] .
 - [11] H. Georgi, (2014), arXiv:1408.1161 [hep-ph] .
 - [12] S.-F. Ge, JHEP **1505**, 066 (2015), arXiv:1408.3823 [hep-ph] .
 - [13] Y. Bai, Z. Han, and R. Lu, JHEP **1503**, 102 (2015), arXiv:1411.3705 [hep-ph] .
 - [14] T. Kaufmann, A. Mukherjee, and W. Vogelsang, Phys.Rev. **D91**, 034001 (2015), arXiv:1412.0298 [hep-ph] .
 - [15] J. Thaler, (2015), arXiv:1506.07876 [hep-ph] .
 - [16] G. C. Fox and S. Wolfram, Phys.Rev.Lett. **41**, 1581 (1978).
 - [17] G. C. Fox and S. Wolfram, Nucl. Phys. **B149**, 413 (1979), [Erratum: Nucl. Phys.B157,543(1979)].
 - [18] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **0804**, 005 (2008), arXiv:0802.1188 [hep-ph] .
 - [19] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), arXiv:0802.1189 [hep-ph] .
 - [20] M. Cacciari, G. P. Salam, and G. Soyez, Eur.Phys.J. **C72**, 1896 (2012), arXiv:1111.6097 [hep-ph] .
 - [21] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, *et al.*, Comput.Phys.Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph] .
 - [22] J. Thaler and K. Van Tilburg, JHEP **03**, 015 (2011), arXiv:1011.2268 [hep-ph] .
 - [23] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, JHEP **08**, 001 (1997), arXiv:hep-ph/9707323 [hep-ph] .
 - [24] M. Wobisch and T. Wengler (1998) arXiv:hep-ph/9907280 [hep-ph] .
 - [25] G. Aad *et al.* (ATLAS), (2015), arXiv:1506.00962 [hep-ex] .
 - [26] G. Altarelli, B. Mele, and M. Ruiz-Altaba, Z. Phys. **C45**, 109 (1989), [Erratum: Z. Phys.C47,676(1990)].
 - [27] Y.-T. Chien, Phys. Rev. **D90**, 054008 (2014), arXiv:1304.5240 [hep-ph] .
 - [28] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, Phys. Rev. Lett. **110**, 212001 (2013), arXiv:1209.2421 [hep-ph] .
 - [29] Y. Cui, Z. Han, and M. D. Schwartz, Phys.Rev. **D83**, 074023 (2011), arXiv:1012.2077 [hep-ph] .
 - [30] G. C. Fox and S. Wolfram, Phys. Lett. **B82**, 134 (1979).